The Ruth H. Hooker Research Library

and Technical Information Center

RETRIEVAL OF OPTICAL IMAGES USING CUADRA STAR AND GENESYS IN THE RUTH H. HOOKER RESEARCH LIBRARY AND TECHNICAL INFORMATION CENTER

Murray L. Bradley
Naval Research Laboratory
Doris R. Folen
Kestrel Associates, Incorporated

- **Keywords:**

  Optical Images, Technical Reports, Cuadra Star, Genesys, Information Retrieval, Naval Research Laboratory

- **Abstract:**

  The Ruth H. Hooker Research Library and Technical Information Center of the Naval Research Laboratory (NRL) in Washington, D. C. is taking advantage of recent technical advances to place the optical images of its scientific reports on disk and store them in an autochanger roughly the size of a standard home refrigerator. This conversion allows the Library to store its report collection in a small fraction of the space once needed, and to retrieve reports almost instantaneously. The full text of over 60,000 reports averaging 55 pages per report (about one-fifth of the report collection), has already been stored as optical images and 36,000 more will be scanned by the end of the current year. These images are on 12-inch optical disks in a 50 platter Sony autochanger. In addition to the autochanger, the basic system consists of an Alphamicro minicomputer, an optical server, a file server, PCs for transferring images to disk, scanners, printers, and workstations. The reports are fully indexed in a Cuadra Star database. Page images of the reports are retrieved using the ImageExtender capability of the GeneSys software. The Library user searches the STAR database from a PC to identify those reports needed. Once the reports are identified a scroll bar is placed over the accession numbers of these reports and with the touch of a key the images are retrieved. Pages of reports can be displayed one at a time on the screen or, in a Windows version, pages from two or more reports can be displayed simultaneously. The same report can be viewed simultaneously at more than one workstation. The user can flip back and forth through the report on the screen, front to back or back to front, or go directly to a certain page. All or part of a report can be printed out by the user on a high speed printer.

1. INTRODUCTION: BACKGROUND AND PROBLEM STATEMENT

The Naval Research Laboratory created the Documents Section of the Ruth H. Hooker Research Library and Technical Information Center in 1945. At that time, the research community assumed that the end of World War II research would mean the end of report production by government agencies and contractors. The Laboratory believed that after the reports archive collection was gathered together and organized, the collection would remain static. However, the opposite happened. The production of technical reports continued to accelerate during the next four decades as scientific and government-sponsored research increased rather than decreased. The Section now contains over 16,000 NRL originated reports; over 600,000 additional titles have been accessioned and cataloged by the Section since its establishment. Over the years this collection of 600,000 titles has been reduced so that 300,000 hard copies remained on the shelves. These remaining reports now are considered invaluable and irreplaceable, dating back to the early 1940's and before. Their subject content is the result of basic, seminal research done in the areas of physics, chemistry, oceanography, electronics, metallurgy, optics, engineering, intelligence, artificial intelligence, etc. They originate from research facilities located at government installations, universities, and corporate laboratories not only in this country but also abroad. In many cases, the reports

represent the only known copy in existence. Concurrently, the "Information Explosion" of the last half-century has produced not only a mammoth increase in the size of technical journals but also has given birth to a flood of new ones. Desperate for space, librarians at the Naval Research Laboratory began to look to media other than paper to store information. The use of microfiche and microfilm helped to alleviate the problem, but only temporarily. The library continued to grow. Solving this space problem is the most visible benefit of optical technology. As the collections grew so did the problems of retrieval. Identifying items needed in this sea of information and actually retrieving them from the shelf started to take longer and longer. Efficient automated search systems assisted in solving the information retrieval problems, but physical retrieval of the reports remained a problem. Maintenance of the growing collection generated another set of problems. The tasks of filing, re-filing, shelf checking and inventory, repair of damaged materials, charge out, etc. became more and more time-consuming. Optical storage solves these problems. Maintenance of items filed as images takes very little time. There is no manual filing or refiling. Charge out and recall are eliminated as items are printed out and given to requesters to keep.

2. THE OPTICAL DISK DECISION

Optical disk technology is not the only answer to the problems that libraries face. The use of optical storage in combination with CD- ROM, microfiche, microfilm and other technologies available today is actually the sensible answer. Optical storage is, in the case of the NRL library, the method of choice to solve only some of the problems and has not eliminated the use of other media. The NRL Library actively uses commercial CD-ROM products which it has put on its campus-wide network and has a growing collection of microfiche and microfilm[1].

2.1 Storage

Space, a major consideration in most libraries, became an acute problem for the NRL library in 1987. The Library lost 3600 square feet previously occupied by a large part of the collection. The reports were going to be moved during a building renovation and there was no convenient available space for them to be relocated. An intensive study and comparison of available technology pointed to optical storage as the solution of choice:

- The cost was less than or comparable to microfilming or mastering CD-ROMs;
- The project could be done in-house at our own pace;
- Once placed digitally on optical disk, images can be downloaded and transferred to other optical systems or other media. For example, the Library is currently upgrading its 3.2 gigabyte disks to 6.5 gigabyte disks.
- 12-inch optical disks provide a storage medium many times greater than CD-ROM. The 12" disks can be conveniently housed in autochangers (In 1987, this option was not available for CD-ROMs). Autochangers can be connected together with one optical server and accessed on a network. In theory, the entire book, journal and report collection of the NRL Library could, with the addition of new autochangers, reside on a single optical disk system and still have plenty of room to grow. Moreover, additional optical servers could be networked should even greater storage capacity be required;
- Optical disks are not "mastered". Images can be placed on a disk and retrieved immediately. Additional images can be added later as desired;
- The system is a "stand-alone" in-house system and does not have to be 100% compatible with anyone else's system;
- Optical disks provide a large amount of storage in a very small area. Items are readily retrievable, regardless of the order in which they are placed on the platter. The use of microfiche, microfilm or even CD-ROM, requires a pre-determination often to store items for ease of recovery.

2.2 Maintenance

As report collections grow and age, problems of maintenance arise. Misfiling and deterioration of pages become major concerns. Changes and corrections to reports are constantly being received and are a bother to incorporate into the original report. Placing the reports on optical disk solved these problems.

- The National Institute of Standards and Technology has estimated a disk life expectancy of 57 to 121 years, depending upon the temperature and humidity used to prematurely age the disks[2];

- Placing reports on disk as optical images provides a better and a more constant quality of image than does microfilming or microfiching;

- The order of reports placed on a disk is immaterial and does not affect retrieval time;

- Changes and/or corrections can be conveniently added later at any time on the optical disk and the changes/corrections automatically linked with the original document so that the distance in storage space between the original and the change/correction is transparent to the user. This cannot be done with microfiche, microfilm or CD-ROM without refilming or re-recording the entire unit.

2.3 Retrieval

Optical disk was considered the best of the new technologies for the retrieval of items because:

- Optical disk autochangers can be placed on a network and images retrieved locally or remotely without the patron or librarian leaving his or her desk;

- Retrieval of optical images takes only a matter of seconds;

- Since the number of items that needed to be stored would require hundreds of thousands of microfiche, thousands of reels of film, or hundreds of CD-ROMs, retrieval of the stored product would be a problem if placed on one of these media. The correct microfiche CD-ROM or film would have to be found, mounted in a player or a reader before the item could be retrieved.

3. THE OPTICAL DISK SYSTEM CONFIGURATION (Figure 1)

A Sony Writable Optical Disk Autochanger Model WDA-610 is the nucleus of the system. It is designed specifically for use with Sony's 12-inch optical disks. One 12-inch "Write Once Read Many" (WORM) optical disk can provide 6.5 gigabytes (GB) of digital data, enough to store the contents of 110,000 typewritten pages (2,000 technical reports). Fifty disks can be housed in one autochanger.

A TDC DocuScan DS-4530 scanner has been incorporated into the system as the primary scanner. This scanner has a sheet feeder, a monitor, and is capable of duplex scanning at a maximum rate of 40 sheets (i.e., two pages) a minute. Scanning is being done at 300 dpi. The scanner compresses the image into the industry standard CCITT Group IV format. After scanning, the data moves across three PCs. First, "Cruncher" changes the TDC scanner output into a Genesys input format. Next, "Mover" transfers the data from the NOVELL file server to the Genesys optical server. The last PC, "Verifier," compares the image on optical disk with the image zon the NOVELL file server and removes it from the magnetic disk when the two images match exactly. The net result of these three operations is the error-free movement of the data onto the optical disk. Custom software interfaced with the above equipment automatically locates and fixes any electronic or image errors.

The software used for storage and retrieval of the images is the Genesys program with its ImageExtender capability for linking the image with any existing database or catalog. Images are recalled at the viewing stations by first searching the STAR database mounted on the Alphamicro minicomputer. The image is retrieved from the file server, decompressed, and either read at a work station or printed out on an HP Laser-Jet printer. The equipment is networked internally on a Ethernet LAN and could be connected to the campus- wide network, which is part of the Internet.

A heavy-duty paper cutter is used to prepare the reports for scanning. Staples and bindings are removed. The cutter is capable of removing staples and bindings in stacks up to three inches thick.

An 18KVA Uninterruptible Power Supply was installed to provide power in the event of a power failure or general blackout.

4. THE CONVERSION PROCESS

The reports to be scanned take a one way trip to the optical platter. The image quality is consistently good enough to allow the destruction of the reports after scanning. Because the entire collection of reports is to be scanned, no particular care has been taken with scanning order. Unlike full size paper reports filed on a shelf, images do not have to be placed on the disk in any particular order.

Use of the TDC DS-4530 scanner is almost fully automatic. Bar coded accession numbers are read by the scanner. The bar code signals the first page of the next document. Quality of the scanned image is checked page by page on a preview monitor by the operator during scanning. TDC's DOCUSCAN software screen monitors the scanning process. Items such as scanning mode (simplex/duplex) and throughput rates are displayed. Statistics generated from this software provide data for management analysis.

5. USING THE SYSTEM

Reports are retrieved by the user at a viewing station. The user searches the on-line catalog. When the search results are shown on the screen the report is recalled by placing the cursor in the "Document Number" field and pressing a function key. The report originally appears on the screen with the search results. The user can view the report this way or the screen can be changed to display only the report.

The report can be viewed page by page on the screen or the user can skim through the report, back and forth, at his or her own pace. A page can be enlarged for detail or reduced in size to allow other pages to be displayed on the screen. All or part of the report can be printed out by a simple command. In a Windows environment, more than one report can be viewed on a screen and the same report can be viewed simultaneously at more than one work station at the same time.

The ability to identify the report or reports to be retrieved is crucial to the process. Thousands of records placed on optical disk can rapidly submerge individual items in a ocean of data. An indexing program to access information is an expensive choice. This program must adequately search the database, yet be flexible enough to change.

The NRL Library finessed this retrieval dilemma by opting to keep the retrieval system entirely separate from the Optical Disk System. The reasoning was:

- An excellent, powerful retrieval database was already in place;
- There was a need for the quickest and the most expedient means of getting the reports to optical disk and a separate retrieval system supported that goal;
- A separate system is less expensive and easier to manage, input can be done at its own pace, and anything on the system can be changed at any time;
- At the beginning of the planning stage the optical disk system was thought of as simply an alternative way of storing reports;
- The Genesys ImageExtender allowed the use of a pre-existing database to match retrieved data with images;
- With the Genesys ImageExtender the fact that the cataloging database is separate from the image file is transparent to the user.

Since August of 1987, the Library has had in place the Cuadra STAR retrieval system. This system provides a very large number of fields so that a report can be indexed in every conceivable way. Indexes such as accession number, title, author, subject, contract, words in an abstract, size, etc. Searching may be done by using any field separately or combined. The fields may be searched in full or by individual words, by subfields or by truncation. Boolean searching of combinations of fields or search results is also possible. The results of all searches can be displayed or printed out by any or all of the fields used in the indexing in any order or in any combination.

The flexibility of the system allows for the addition of a field or subfield as needed. For example, in 1992, the

Library added an "optical disc" field to indicate those reports already on disc. Global macros allow for sweeping changes in indexing if necessary. Authority lists and lookup tables are used for controlling subject and descriptive cataloging terms, but fields are also available for the cataloger to use for new terms.

6. THE OPTICAL DISC SYSTEM AS PART OF THE LIBRARY

The optical disc system is being integrated into the overall operation of the Library. A portion of the images are being transferred to a second optical system in another section of the Library where they will be offered in combination with full text images of reprints and journals articles. This optical system will be networked in the Library where patrons can view images at their leisure as they would books and journals in study carrels. Then, in the near future, this system will be made available on the NRL campus network and the scientists will be able to call up information in the convenience of their own offices.

7. SUMMARY AND CONCLUSION.

Optical image technology offers vast possibilities to restructure the traditional operation of a library and to allow libraries to change the role they have previously played in the use of information. Because so much information can be placed conveniently in a very small space and quickly retrieved, libraries can become the focal point for data storage in an organization: information such as archive copies of incoming and outgoing mail, administrative reports, budget and personnel records and historical documents, all of which the library can store and retrieve as needed - a service that most libraries have not had the capability of providing in the past. Networking of images will suggest additional roles for libraries as these images can be sent or called up and printed out in offices anywhere in the organization. Thus, the concept of the "Virtual Library" has come of age. Technology such as optical systems has provided the impetus for this very important new aspect of the library world.

8. REFERENCES

1. Laurie E. Stackpole, Roderick D. Atkinson, and John Yokley. "Campus-Wide Network Access to CD-ROM Databases", *Proceedings of the Thirteenth National Online Meeting*, 1992. Pages 379-385.

2. Podio, Fernando. Development of a Testing Methodology to Predict Optical Disk Life Expectancy Values. (NIST Special Publication 500-200). Washington: U. S. Government Printing Office, 1991. Page 28.

---

webmaster@library.nrl.navy.mil

Updated: 27-NOV-95

Maintained by: Fred Rettenmaier